



# UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE  
United States Patent and Trademark Office  
Address: COMMISSIONER FOR PATENTS  
P.O. Box 1450  
Alexandria, Virginia 22313-1450  
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
09/607,710	06/30/2000	Marc Alexander Najork	18449772-0288-999	8784

7590 09/22/2003

GARY S. WILLIAMS  
PENNIE & EDMONDS LLP  
3300 HILLVIEW AVENUE  
PALO ALTO, CA 94304

EXAMINER

OSMAN, RAMY M

ART UNIT

PAPER NUMBER

2157

6

DATE MAILED: 09/22/2003

Please find below and/or attached an Office communication concerning this application or proceeding.

**Office Action Summary**

Application No.

09/607,710

Applicant(s)

NAJORK ET AL.

Examiner

Ramy M Osman

Art Unit

2157

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

**Period for Reply**

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If the period for reply specified above is less than thirty (30) days, a reply within the statutory minimum of thirty (30) days will be considered timely.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133).
- Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

**Status**

- 1) ☐ Responsive to communication(s) filed on \_\_\_\_.
- 2a) ☐ This action is FINAL. 2b) ☒ This action is non-final.
- 3) ☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

**Disposition of Claims**

- 4) ☒ Claim(s) 1-52 is/are pending in the application.
- 4a) Of the above claim(s) \_\_\_\_ is/are withdrawn from consideration.
- 5) ☐ Claim(s) \_\_\_\_ is/are allowed.
- 6) ☒ Claim(s) 1-52 is/are rejected.
- 7) ☒ Claim(s) 20 is/are objected to.
- 8) ☐ Claim(s) \_\_\_\_ are subject to restriction and/or election requirement.

**Application Papers**

- 9) ☐ The specification is objected to by the Examiner.
- 10) ☒ The drawing(s) filed on 6/30/00 is/are: a) ☐ accepted or b) ☒ objected to by the Examiner.
- Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).
- 11) ☐ The proposed drawing correction filed on \_\_\_\_ is: a) ☐ approved b) ☐ disapproved by the Examiner.
- If approved, corrected drawings are required in reply to this Office action.
- 12) ☐ The oath or declaration is objected to by the Examiner.

**Priority under 35 U.S.C. §§ 119 and 120**

- 13) ☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).
- a) ☐ All b) ☐ Some \* c) ☐ None of:
1. ☐ Certified copies of the priority documents have been received.
2. ☐ Certified copies of the priority documents have been received in Application No. \_\_\_\_.
3. ☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).
- \* See the attached detailed Office action for a list of the certified copies not received.
- 14) ☐ Acknowledgment is made of a claim for domestic priority under 35 U.S.C. § 119(e) (to a provisional application).
- a) ☐ The translation of the foreign language provisional application has been received.
- 15) ☐ Acknowledgment is made of a claim for domestic priority under 35 U.S.C. §§ 120 and/or 121.

**Attachment(s)**

- 1) ☒ Notice of References Cited (PTO-892) 4) ☐ Interview Summary (PTO-413) Paper No(s). \_\_\_\_.
- 2) ☐ Notice of Draftsperson's Patent Drawing Review (PTO-948) 5) ☐ Notice of Informal Patent Application (PTO-152)
- 3) ☒ Information Disclosure Statement(s) (PTO-1449) Paper No(s) \_\_\_\_.
- 6) ☐ Other: \_\_\_\_\_

## **DETAILED ACTION**

### ***Drawings***

1. The drawings are objected to because reference number 141, mentioned on page 7 line 15, is missing from figure 2. A proposed drawing correction or corrected drawings are required in reply to the Office action to avoid abandonment of the application. The objection to the drawings will not be held in abeyance.

### ***Claim Objections***

2. Claim 20 objected to because of the following informalities: "step (e1)" should be changed to "step (d1)". Appropriate correction is required.

### ***Claim Rejections - 35 USC § 102***

3. The following is a quotation of the appropriate paragraphs of 35 U.S.C. 102 that form the basis for the rejections under this section made in this Office action:

(a) the invention was known or used by others in this country, or patented or described in a printed publication in this or a foreign country, before the invention thereof by the applicant for a patent.

4. Claims 1-4,7-10,13-17,20-25,28-33,36-40,43-48,51 and 52 are rejected under 35 U.S.C. 102(a) as being anticipated by Monier (U.S. Patent No. 5,974,455).

Monier teaches the invention as claimed including a method, a computer program and a web crawler system for efficient representation of data set addresses in a web crawler (see Monier, Abstract).

5. In reference to claims 1,13,22,23,31,38 and 46, Monier teaches downloading data sets from among a plurality of host computers comprising the following steps:

Storing representations of data set addresses in a set of data structures, including a first buffer, a second buffer and a first disk file, wherein representations of data set addresses stored in the first disk file are ordered (column 3, lines 1-35, Monier discloses storing URL representations in a set of data structures, including a hash table (stored in random access memory (RAM)), an append buffer (stored in RAM) and a sequential disk file, wherein the representations are stored sequentially in the disk file).

Selecting as a current buffer one of the first and second buffers (column 6, lines 35-45, Monier discloses selecting and managing a current buffer among the hash table and append buffer).

Downloading at least one data set that includes addresses of one or more referred data sets (column 5, lines 20-30, Monier discloses fetching web pages that include URL's of one or more referred web pages).

Identifying the addresses of the one or more referred data sets (column 5, lines 20-30, Monier discloses analyzing and identifying the addresses of the one or more referred web pages).

For each identified address:  
Generating a representation of the identified address (column 5 line 55 – column 6 line 22, Monier discloses generating a fingerprint representation of the specified URL), and  
Determining whether the representation is stored in the buffer, and when this determination is negative, storing the representation in the buffer (column 5 line 43 – column 6 line 22, Monier

discloses determining whether the representation is stored in the hash table, and when this determination is negative, storing the representation in the hash table).

When the buffer reaches a predefined full condition:

Ordering the contents of the buffer according to the representations (column 6, lines 1-33, Monier discloses ordering the contents of the hash table according to the fingerprint representations), and

Performing an ordered merge of the contents of the buffer into the contents of the first disk file (column 6 line 22 – column 7 line 12, Monier discloses performing a merge of the contents of the hash table into the contents of the disk file), and

Selecting the other buffer as the current buffer, wherein the previously current buffer is identified as a non-current buffer (column 6, lines 22-67, Monier discloses selecting the append buffer as the current buffer, wherein the hash table is identified as a non-current buffer).

6. In reference to claims 2,14,24,32,39 and 47, Monier teaches the method, the computer program and the web crawler system of claims 1,13,22,23,31,38 and 46 above, wherein after determining that the representation is not stored in the buffer, the identified address is stored in the buffer (column 5 line 43 – column 6 line 22, Monier discloses that after determining that the representation is not stored in the hash table, the identified address is stored in the hash table).

7. In reference to claims 3,15,25,33,40,48, Monier teaches the method, the computer program and the web crawler system of claims 1,13,23,31,38 and 46 above, wherein:  
After determining that the representation is not stored in the buffer, the identified address is stored in a second disk file (column 9, lines 25-40, Monier discloses after determining that the

representation is not stored in the hash table, the identified address is stored in a second disk file), and

Additionally storing with each representation in the buffer a pointer to the corresponding address stored in the second disk file (column 3, lines 1-20, column 5 & column 6, lines 20-53, Monier discloses additionally storing with each representation in the RAM a pointer to the corresponding address stored in the second disk file), and

While ordering the contents of the buffer, keeping with each representation in the buffer its pointer to the corresponding address in the second disk file (column 5 & column 6, lines 20-53, Monier discloses while ordering the contents of the hash table (in RAM), keeping with each representation in the hash table its pointer to the corresponding address in the disk file).

8. In reference to claims 4 and 16, Monier teaches the method of claims 3 and 15 above, wherein when the buffer reaches a predefined full condition:

Each representation in the buffer stores an associated flag, setting the flag to a first value when the representation is equal to a representation previously stored in the first disk file, and setting the flag to a second value, when the representation is not equal to any representation previously stored in the first disk file (column 5 lines 25-35, & column 8 lines 45-65, Monier discloses each representation in the hash table stores an associated “ fetched flag”, setting the flag to a first value when the representation is equal to a representation previously stored in the disk file, and setting the flag to a second value, when the representation is not equal to any representation previously stored in the disk file), and

Each representation whose flag is set to the second value, scheduling the corresponding data set for downloading (column 9, lines 25-50, Monier discloses each representation whose flag is set

Art Unit: 2157

to the second value and marked as “not fetched”, scheduling the corresponding data set for fetching).

9. In reference to claims 7,20,28,36,43 and 51, Monier teaches the method, the computer program and the web crawler system of claims 1,13,23,31,38 and 46 above, wherein the representation of the identified address comprises a checksum of at least a portion of the identified address (column 5 line 55 – column 6 line 22, Monier discloses the representation of the identified URL comprising a fingerprint of at least a portion of the identified URL).

10. In reference to claims 8,21,29 and 44, Monier teaches the method, the computer program and the web crawler system of claims 1,13,23 and 38 above, wherein:

Determining whether the representation is stored in a cache before determining whether the representation is stored in the buffer (columns 6&7, Monier discloses determining whether the representation is stored in append buffer before determining whether the representation is stored in an input buffer (in RAM)), and

Determining whether the representation is stored in a cache, and if positive, skipping the determination of whether the representation is stored in the buffer (columns 6&7, Monier discloses determining whether the representation is stored in an append buffer, and if positive, skipping the determination of whether the representation is stored in the input buffer), and

When the representation is not stored in the cache, the cache has not reached a predefined full condition, and other predefined criteria are met, adding the representation to the cache (columns 6&7, Monier discloses when the representation is not stored in the append buffer, the host name table has not reached a predefined full condition, and other predefined criteria are met, adding the representation to the input buffer), and

Art Unit: 2157

When the representation is not stored in the cache, the cache has reached said predefined full condition, and said other predefined criteria are met, evicting a stored representation from the cache in accordance with an eviction policy and adding the representation to the cache (columns 6&7, Monier discloses when the representation is not stored in the append buffer, the append buffer has reached said predefined full condition, and said other predefined criteria are met, evicting a stored representation from the append buffer in accordance with an eviction policy and adding the representation to the append buffer).

11. In reference to claims 9,10,17,30,37,45 and 52, Monier teaches the method, the computer program and the web crawler system of claims 1,23,31,38 and 46 above, wherein when a representation in the first buffer is not found in the first disk file during merging, scheduling the corresponding data set for downloading (columns 6-8, Monier discloses that when a representation in the hash table is not found in the disk file during merging, scheduling the corresponding web page for fetching).

### ***Claim Rejections - 35 USC § 103***

12. The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

(a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negated by the manner in which the invention was made.



Art Unit: 2157

13. Claims 5,6,11,12,18,19,26,27,34,35,31,32,49 and 50 are rejected under 35 U.S.C. 103(a) as being unpatentable over Monier (U.S. Patent No. 5,974,455) in view of Cabrera et al. (U.S. Patent No. 5,953,729).

14. In reference to claims 5,11,18,26,34,41 and 49, Monier teaches the method, the computer program and the web crawler system of claims 1,13,23,31,38 and 46 above.

Monier does not teach storing representations of data set addresses in a sparse disk file which is divided into portions (or sub-files), each portion having a starting address and contents comprising an ordered list of representations of data addresses. However, Cabrera teaches sparse file technology divided into clusters each having a cluster number (column 9, lines 40-66).

It would have been obvious to one having ordinary skill in the art to modify Monier by storing URL representations in a sparse file as per the teachings of Cabrera so as to minimize the overhead in managing and ordering the contents on the disk file.

15. Monier does not teach merging the contents of the buffer with the ordered contents of the sparse disk file to include determining a starting address for a corresponding portion of the sparse disk file. However, Cabrera teaches sparse file technology which can indicate starting cluster numbers for portions of the sparse file (columns 9&10).

It would have been obvious to one having ordinary skill in the art to modify Monier by when merging the contents of the hash table with the ordered contents of the sparse file, to include determining a starting cluster number for a corresponding portion of the sparse disk file as per the teachings of Cabrera so as to minimize the overhead for merging and ordering of the contents on the disk file.

16. Monier does not teach merging the contents of the buffer with the ordered contents of the sparse disk file to include performing an ordered merge of a subset of the buffer, starting at the representation for which the starting address was obtained, into the contents of the corresponding portion. However, Cabrera teaches sparse file technology which can indicate starting cluster numbers for portions of the sparse file (columns 9&10).

It would have been obvious to one having ordinary skill in the art to modify Monier by when merging the contents of the hash table with the ordered contents of the sparse disk file to include performing an ordered merge of a subset of the hash table, starting at the representation for which the starting address was obtained, into the contents of the corresponding portion as per the teachings of Cabrera so as to minimize the overhead in merging and ordering the contents on the disk file.

17. In reference to claims 6,12,19,27,35,42 and 50, Monier teaches the method, the computer program and the web crawler system of claims 1,13,23,31,38 and 46 above.

18. Monier does not teach storing representations of data set addresses in a sparse disk file having empty entries interspersed among entries storing said representations. However, Cabrera teaches sparse file technology which comprises a mixture of zero data and non-zero data (column 7, lines 20-50).

It would have been obvious to one having ordinary skill in the art to modify Monier by storing representations of data set addresses in a sparse disk file having zero data interspersed among data of said representations as per the teachings of Cabrera so as to minimize the overhead in sequentially ordering the data contents on the disk file.

19. Monier teaches sequentially scanning the disk file via an input buffer, starting at the representation for which a starting address was obtained, until a representation matching the respective representation is found (column 6 lines 35-67 & column 9 lines 25-50). Monier does not teach scanning the disk file until one of the empty entries is found, and when an empty entry is found storing the respective representation in the empty entry. However, Cabrera teaches sparse file technology which comprises a mixture of zero data and non-zero data (column 7, lines 20-50).

It would have been obvious to one having ordinary skill in the art to modify Monier by scanning the disk file until one of the zero data entries is found as per the teachings of Cabrera, and when zero data entry is found storing the respective representation in the zero data entry, so as to minimize the overhead of ordering the data contents on the disk file while merging the contents of the hash table with the contents of the disk file.

### ***Conclusion***

20. The prior art made of record and not relied upon is considered pertinent to applicant's disclosure.

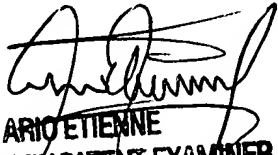
- US Patent No. 6,547,829 B1
- US Patent No. 5,913,208 A
- US Patent No. 5,564,037 A
- US Patent No. 5,893,086 A
- US Patent No. 6,490,658 B1

Any inquiry concerning this communication or earlier communications from the examiner should be directed to Ramy M Osman whose telephone number is (703) 305-8050. The examiner can normally be reached on Monday through Friday 9AM to 5PM.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Ario Etienne can be reached on (703) 305-7562. The fax phone number for the organization where this application or proceeding is assigned is (703) 872-9306.

Any inquiry of a general nature or relating to the status of this application or proceeding should be directed to the receptionist whose telephone number is (703) 305-9600.

RMO  
August 28, 2003

  
ARIO ETIENNE  
SUPERVISORY PATENT EXAMINER  
TECHNOLOGY CENTER 2100